



Article Title: Students Performance Prediction and Classification Using Ridge Regression Algorithm

Students Performance Prediction and Classification Using Ridge Regression Algorithm

Pooja. S. S ¹, Dr. Gladrene Sheena Basil ²,

¹Student, ²Associate Professor,
Department of Computer Science and Engineering, Loyola Institute of Technology
and Science,

¹sspooja236@gmail.com, ² gladrene.basil@gmail.com

ABSTRACT

Education system is very old and due to a large population of students, there are some serious issues in analyzing and predicting students' performance. Every institution has its own set of standards for evaluating student success. In existing system there is no proper procedure for monitoring and analyzing a student's performance and progress. Here the focus is given to additional external factors like students status, parent occupation etc.. That will be more effective in visualizing and analyzing student's performance. This project proposed Ridge regression Algorithm is used.

Keywords: Ridge Regression, Predication, Accuracy, Visualizing, Analyzing, Performance.

1 Introduction

The main goal of every university is to provide quality knowledge and skills to students so that they are competitive in the labour market. One way to achieve this goal is to timely predict student performance. Thus, it is easy to identify students who need support and take measures to improve educational outcomes. This would help teachers to provide an effective approach to teaching. The analysis of students' progress during their studies provides the university management with information about the probability of success of each student. Traditionally, this analysis is done by the lecturers who use their interactions with students in classroom activities and mid-term assessments to identify those at risk of dropping out and take timely action. The objective of this paper is to predict students' performance based on several inputs including academic and non-academics factors. For this, Ridge Regression Algorithm is used.

2 Related Works

There are many studies in the literature related to the students' performance using machine learning. Most researchers in this area seek to identify the most appropriate algorithm by which predictions can be made, and to identify the features that can be used for forecasting. KNN, BNN, ANN and random forest multiple ML algorithms have been used better planning and decision making to enrich quality of education. [2] Performance of multiple supervised ML algorithms are checked on two academic years' student dataset like academic,



Article Title: Students Performance Prediction and Classification Using Ridge Regression Algorithm

behavioural and demographic factors.

Predicting students' performance [3] helps to design effective mechanisms to improve the results of an institution and to avoid dropout. Through an extensive survey of previous researches it is found that SVM is highly efficient and the prediction is accurate.

In [4] Back Propagation (BP), SVM and other ML algorithms have also been applied in the year 2019 to predict students' performance on two different data sets. Mean Square Error, EV score was taken as metrics to measure the performance of the algorithms which show that BP algorithm gave better result.

Features that can be used to make predictions should be identified, many may be interrelated. When the prediction of the performance is effective, the instructors can allocate resources in an efficient manner and instructions can be given clearly.

The qualities [6] and abilities are assessed by the examinations with the aim of finding out what the students 'can do' instead of 'what they know'. The independent variable IQ has a significant effect on the examination performance.

Two prediction models [7], KNN and SVM were used for forecasting the performance of students. Such forecasting is helpful for taking precautions, early detection of trouble and taking instant action.

Predicting the potential for dropout is necessary to prevent dropout and to suggest a workable solution for completing the course without drop outs. [8] Artificial Neural Network (ANN), Decision Tree (DT) and Bayesian Network were used to develop the prediction model.

Large volume of data in the educational databases is a challenge to evaluate the performance of the students [9]. A prediction algorithm was suggested to identify the most important attributes in the data of students. Neural Networks and Decision Tree were used to conduct the meta-analysis so as to predict the performance of students.

There is an explosive growth of data in educational institutions. Prediction of academic results will help to improve the quality of managerial decisions [10]. Smooth Support Vector Machine and Kernel k-means clustering were used for this purpose.

In e-learning, measuring student's performance [11] is all the more difficult. Synthetic Minority Oversampling Technique and random sample replication is used are used to predict the students' performance and to recommend a system to improve weaker students

3 Research Methodology

The methodology adopted in this paper comprises of four stages: Data Acquisition, Data Preprocessing and Feature Selection, Data Visualization, Classification and Prediction. The prediction class can be one of the following: (i) higher chances of getting pass and (ii) lower chances of getting pass.



Article Title: Students Performance Prediction and Classification Using Ridge Regression Algorithm

3.1 Data Acquisition

This dataset containing data collected from Kaggle Students Performance in Exams.

Figure 1 shows data set

	gender	race_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score	reading_score	writing_score
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75

Figure 1: Dataset

3.2 Data Preprocessing and Feature Selection

After taking data, pre-processing is done to check whether it contains any missing values, Duplicates and Check the number of Unique values of each column. Check statistics of data set. Check various categories present in the different categorical column.

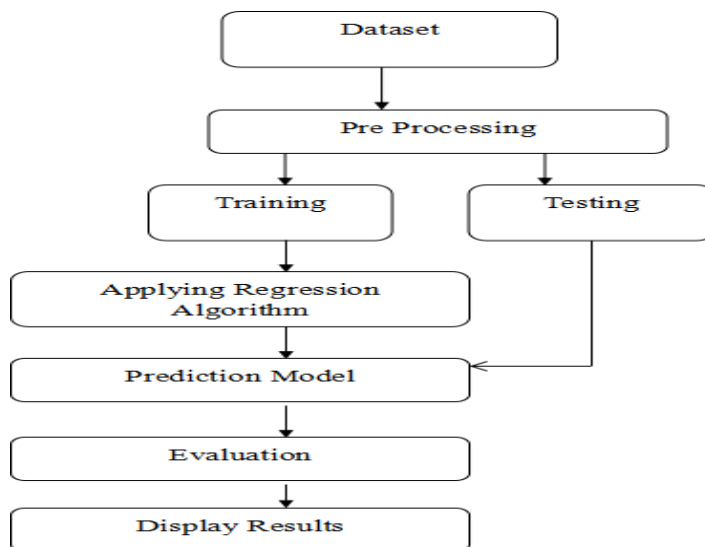


Figure 2: flowchart

3.3 Ridge Regression

Build the Ridge Regression model. And then apply the fitting process for the train set. Ridge Regression model takes some parameters. The main purpose to find the coefficients that minimize the sum of error squares by applying a penalty to these coefficients.

3.4 Prediction

The regression techniques were applied after the feature selection process. The Ridge Regression classifier was found to be the best classifier for building our model, outperforming other classifiers in accuracy, precision, recall, and f-score.



Article Title: Students Performance Prediction and Classification Using Ridge Regression Algorithm

3.5 Ridge Regression Algorithm

Ridge regression is one of the types of linear regression in which a small amount of bias is introduced so that we can get better long-term predictions. Ridge regression is a regularization technique, which is used to reduce the complexity of the model. It is also called as L2 regularization.

In this technique, the cost function is altered by adding the penalty term to it. The amount of bias added to the model is called Ridge Regression penalty. We can calculate it by multiplying with the lambda to the squared weight of each individual feature.

$$\sum_{i=1}^M (y_i - y'_i)^2 = \sum_{i=1}^M \left(y_i - \sum_{j=0}^n \beta_j * x_{ij} \right)^2 + \lambda \sum_{j=0}^n \beta_j^2$$

In the above equation, the penalty term regularizes the coefficients of the model, and hence ridge regression reduces the amplitudes of the coefficients that decreases the complexity of the model. The values of λ tend to zero, the equation becomes the cost function of the linear regression model. Hence, for the minimum value of λ , the model will resemble the linear regression model. A general linear or polynomial regression will fail if there is high collinearity between the independent variables, so to solve such problems, Ridge regression can be used. It helps to solve the problems if we have more parameters than samples. Ridge regression is mostly used to reduce the overfitting in the model, and it includes all the features present in the model. It reduces the complexity of the model by shrinking the coefficients.

4 Existing system

In existing system used various analytical methodologies like K Nearest Neighbors (KNN), Logistic Regression to predict student achievement where most of the researchers have used grade points as the assessment process. There is no proper procedure for monitoring and analyzing a student's performance and progress. Difficult to determine the optimal prediction methodology for visualizing student academic growth and performance and affect students' academic performance and achievement.

5 Proposed system

In this project proposed, some external factors along with academic grades are also considered for predicting student performance such as student's gender, parent's education, student's scores etc. Using Ridge Regression (L2 regularization) performance analysis is high accuracy level of results as well as with some existing work.



Article Title: Students Performance Prediction and Classification Using Ridge Regression Algorithm

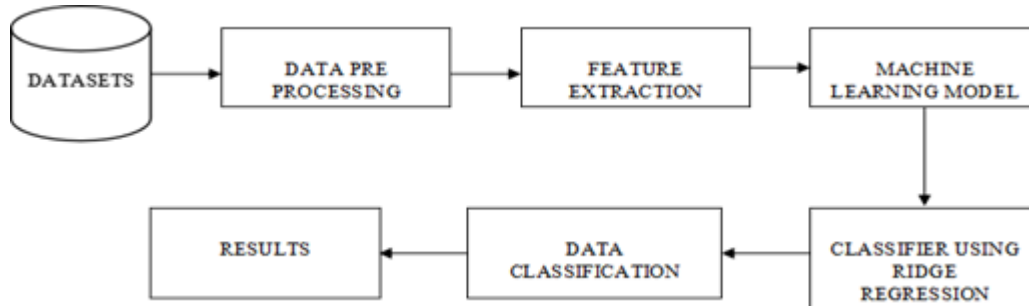


Figure 3: *Architecture diagram*

6 Results and Discussion

In this section, In Existing analysis was performed and it was found that Logistic Regression algorithm is better when compared KNN classifier. By applying Logistic Regression, 86.5% of accuracy was achieved, whereas by applying K-Nearest Neighbors on dataset 78.15% of accuracy was achieved. In proposed Ridge Algorithm obtain 88% high accuracy when compared to KNN and Logistic Regression Algorithm. Figure 4 shows the comparison of the models.

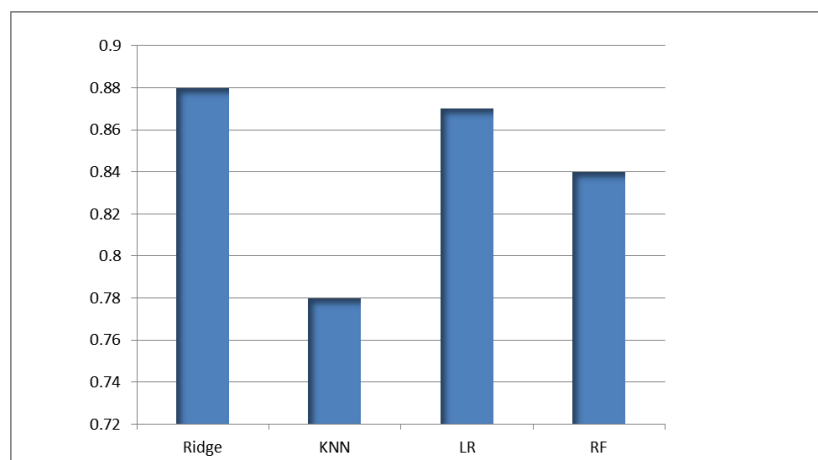


Figure 4: *Bar chart for comparison-based evaluation*

7 Conclusion

Proposed model is also compared with some existing models to show the efficiency of the proposed work. This comparison proves that proposed model can be used efficiently to examines student performance by extracting more insights from the data, as well as to start introducing some beneficial monitoring systems that educationalists can use to precisely assign grades and assist them in improving their students' performance. Student's Performance is related with lunch, race, parental level education. Females lead in pass percentage and also are top-scorers. Student's Performance is not much related with test preparation course finishing preparation course is beneficial.



Article Title: Students Performance Prediction and Classification Using Ridge Regression Algorithm

8 Future Scope

Factors affecting the performance of the students are numerous. In the future work, additional factors should be identified and can be considered. Deep learning algorithms can also be applied to the proposed model in the future, for student performance prediction where researchers can work to provide more accurate results in less computation time.

References

1. E. M. Onyema; K. K. Almuzaini; F. U. Onu; D. Verma; U. S. Gregory; M. Puttaramaiah; R. K. Afriyie, Year: 2022, "Prospects and Challenges of Using Machine Learning for Academic Forecasting," *Computational Intelligence and Neuroscience*, Vol: 2022, pp. 1-7.
2. A. Tarik; H. Aissa; F. Yousef, Year: 2021, "Artificial intelligence and machine learning to predict student performance during the COVID19," in 3rd International Workshop on Big Data and Business Intelligence (BDBI 2021), *Procedia Computer Science* Vol: 184, pp. 835- 84, Poland.
3. A. S. Hashim; W. A. Awadh; A. K. Hamoud, Year: 2020, "Student performance prediction model based on supervised machine learning algorithms," in *IOP Conference Series: Materials Science and Engineering*, Vol: 928, no: 3, IOP Publishing, pp. 032019.
4. Rastrollo-Guerrero J; Gómez-Pulido J; Durán-Domínguez, Year: 2020, Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review. *Applied Science*, Vol: 10, pp. 1042.
5. A. Namoun; A. Alshanqiti, Year: 2020, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," *Applied Sciences*, Vol: 11, no: 1, pp. 237-265.
6. C. Fischer; Z. A. Pardos; R. S. Baker; J. J. Williams; P. Smyth; R. Yu; S. Slater; R. Baker; M. Warschauer, Year: 2020, "Mining big data in education: Affordances and challenges," *Review of Research in Education*, Vol: 44, no: 1, pp. 130-160.
7. G. A. Chong; M. B. Jian; A. Zj, Year: 2020, "Artificial intelligence innovation in education: A twenty-year data-driven historical analysis-ScienceDirect," *International Journal of Innovation Studies*, Vol: 4, no: 4, pp. 134-147.
8. R. Hasan, S. Palaniappan; S. Mahmood; A. Abbas; K. U. Sarker; M. U. Sattar, Year: 2020, "Predicting student performance in higher educational institutions using video learning analytics and data mining techniques," *Applied Sciences*, Vol: 10, no: 11, pp. 3894-3914.
9. A. S. Hashim; W. A. Awadh; A. K. Hamoud, Year: 2020, "Student performance prediction model based on supervised machine learning algorithms," in *IOP Conference Series: Materials Science and Engineering*, Vol: 928, no: 3, IOP Publishing, pp. 032019.



Article Title: Students Performance Prediction and Classification Using Ridge Regression Algorithm

10. B.Sekeroglu; K Dimililer; K Tuncal, Year: 2019, "Student Performance Prediction and Classification using Machine Learning Algorithms", in 8th International Conference on Educational and Information Technology, pp. 7-11.
11. A. Hellas; P. Ihanola; A. Petersen; V. V. Ajanovski; M. Gutica; T. Hynninen; A. Knutas; J. Leinonen, C. Messom; S. N. Liao, Year: 2018, "Predicting academic performance: a systematic literature review," in 23rd annual ACM conference on innovation and technology in computer science education, pp. 175-199.
12. L. Mortada; J. Bolbol; S. Kadry, Year: 2018, "Factors Affecting Students' Performance a Case of Private Colleges in Lebanon," Journal of Mathematical and Statistical Analysis, Vol: 1, no: 1, pp. 105-110.
13. H. Al-Shehri; A. Al-Qarni; L. Al-Saati; A. Batoaq; H. Badukhen; S. Alrashed; J. Alhiyafi; S. O. Olatunji, Year: 2017, "Student performance prediction using support vector machine and k-nearest neighbor," in 30th Canadian conference on electrical and computer, IEEE, pp. 1-4.
14. P. Chaudhury; S. Mishra; H. K. Tripathy; B. Kishore, Year: 2016, "Enhancing the capabilities of student result prediction system," in Second International Conference on Information and Communication Technology for Competitive Strategies, pp. 1-6, 2016.
15. Tan M; Shao P, Year: 2015, "Prediction of student dropout in E-learning program through the use of machine learning method", International journal of emerging technologies in learning, Vol: 10 pp. 11-17.
16. A. M. Shahiri; W. Husain; N. A. Rashid, Year: 2015, "A Review on predicting student's performance using datamining techniques," in Procedia Computer Science, Vol: 72, pp. 414-422.
17. E. M. Onyema; K. K. Almuzaini; F. U. Onu, D. Verma; U. S. Gregory; M. Puttaramaiah; R. K. Afriyie, Year: 2022, "Prospects and Challenges of Using Machine Learning for Academic Forecasting," Computational Intelligence and Neuroscience, Vol: 2022, pp. 1-7.
18. S. Sembiring; M. Zarlis; D. Hartama; S. Ramliana; E. Wani, Year: 2011, "Prediction of student academic performance by an application of data mining techniques," in International Conference on Management and Artificial Intelligence IPEDR, Vol: 6, no: 1.
19. D. Kabakchieva; K. Stefanova; V. Kisimov, Year: 2010, "Analyzing university data for determining student profiles and predicting performance," in In Educational Data Mining.
20. C. Fischer; Z. A. Pardos; R. S. Baker; J. J. Williams; P. Smyth; R. Yu; S. Slater; R. Baker; M. Warschauer, Year: 2020, "Mining big data in education: Affordances and challenges," Review of Research in Education, Vol: 44, no: 1, pp. 130-160.



Article Title: Students Performance Prediction and Classification Using Ridge Regression Algorithm

21. G. A. Chong; M. B. Jian; A. Zj, Year: 2020, "Artificial intelligence innovation in education: A twenty-year data-driven historical analysis-ScienceDirect," *International Journal of Innovation Studies*, Vol: 4, no: 4, pp. 134-147.
22. R. Hasan; S. Palaniappan; S. Mahmood; A. Abbas; K. U. Sarker; M. U. Sattar, Year: 2020, "Predicting student performance in higher educational institutions using video learning analytics and data mining techniques," *Applied Sciences*, Vol: 10, no: 11, pp. 3894-3914.
23. A. S. Hashim; W. A. Awadh; A. K. Hamoud; Year: 2020, "Student performance prediction model based on supervised machine learning algorithms," in *IOP Conference Series: Materials Science and Engineering*, Vol: 928, no: 3, IOP Publishing, pp. 032019.
24. B. Sekeroglu; K. Dimililer; K. Tuncal, Year: 2019, "Student Performance Prediction and Classification using Machine Learning Algorithms", in *8th International Conference on Educational and Information Technology*, pp. 7-11.
25. A. Hellas; P. Ihtola; A. Petersen; V. V. Ajanovski; M. Gutica; T. Hynninen; A. Knutas; J. Leinonen; C. Messom; S. N. Liao, Year: 2018, "Predicting academic performance: a systematic literature review," in *23rd annual ACM conference on innovation and technology in computer science education*, pp. 175-199.
26. L. Mortada; J. Bolbol; S. Kadry, Year: 2018, "Factors Affecting Students' Performance a Case of Private Colleges in Lebanon," *Journal of Mathematical and Statistical Analysis*, Vol: 1, no: 1, pp. 105-110.
27. H. Al-Shehri; A. Al-Qarni; L. Al-Saati; A. Batoaq; H. Badukhen; S. Alrashed; J. Alhiyafi; S. O. Olatunji, Year: 2017, "Student performance prediction using support vector machine and k-nearest neighbor," in *30th Canadian conference on electrical and computer*, IEEE, pp. 1-4.
28. P. Chaudhury; S. Mishra; H. K. Tripathy; B. Kishore, Year: 2016, "Enhancing the capabilities of student result prediction system," in *Second International Conference on Information and Communication Technology for Competitive Strategies*, pp. 1-6.
29. Tan M; Shao P, Year: 2015, "Prediction of student dropout in E-learning program through the use of machine learning method", *International journal of emerging technologies in learning*, Vol: 10 pp. 11–17.
30. A. M. Shahiri; W. Husain; N. A. Rashid, Year: 2015, "A Review on predicting student's performance using datamining techniques," in *Procedia Computer Science*, Vol: 72, pp. 414-422.
31. E. M. Onyema; K. K. Almuzaini, F. U. Onu; D. Verma; U. S. Gregory; M. Puttaramaiah; R. K. Afriyie, Year: 2022, "Prospects and Challenges of Using Machine Learning for Academic



Article Title: Students Performance Prediction and Classification Using Ridge Regression Algorithm

Forecasting," Computational Intelligence and Neuroscience, Vol: 2022, pp. 1-7.

32. S. Sembiring; M. Zarlis; D. Hartama; S. Ramlina; E. Wani, Year: 2011, "Prediction of student academic performance by an application of data mining techniques," in International Conference on Management and Artificial Intelligence IPEDR, Vol: 6, no: 1.

33. D. Kabakchieva; K. Stefanova; V. Kisimov, Year: 2010, "Analyzing university data for determining student profiles and predicting performance," in In Educational Data Mining.